Vos, Murphy, Howard and Daffern, (2017). Determining the Relevant Criteria for Three-dimensional Vocal Tract Characterization, Journal of Voice, http://dx.doi.org/10.1016/j.jvoice.2017.04.001 published online: 21st June 2017.

Determining the Relevant Criteria for Three-dimensional Vocal Tract Characterization

Vos, R.R.(1), Murphy, D.T.(1), Howard, D.M.(2), and Daffern, H.(1)
(1) Department of Electronics, University of York, UK
(2) Department of Electronic Engineering, Royal Holloway, University of London, UK

ABSTRACT
Objective: Soprano singers face a number of specific challenges when singing vowels at high frequencies, due to the wide spacing of harmonics in the voice source. The varied and complex techniques used to overcome these are still not fully understood. Magnetic resonance imaging (MRI) has become increasingly popular in recent years for singing voice analysis. This study proposes a new protocol using three-dimensional MRI to investigate the articulatory parameters relevant to *resonance tuning*, a technique whereby singers alter their vocal tract to shift its resonances nearer to a voice source harmonic, increasing the amplitude of the sound produced.

Methods: The protocol was tested with a single soprano opera singer. Drawing on previous MRI studies, articulatory measurements from three-dimensional MRI images were compared to vocal tract resonances measured directly using broadband noise excitation. The suitability of the protocol was assessed using statistical analysis.

Results: No clear linear relationships were apparent between articulatory characteristics and vocal tract resonances. The results were highly vowel dependent, showing different patterns of resonance tuning and interactions between variables. This potentially indicates a complex interaction between the vocal tract and sung vowels in soprano voices, meriting further investigation.

Conclusions: The effective interpretation of MRI data is essential for a deeper understanding of soprano voice production and, in particular, the phenomenon of resonance tuning. This paper presents a new protocol that contributes toward this aim, and the results suggest that a more vowel-specific approach is necessary in the wider investigation of resonance tuning in female voices.

KEYWORDS: resonance, soprano, tuning, MRI, voice

INTRODUCTION

Resonance tuning is a technique employed by professional soprano singers (although not exclusively [1]) whereby the singers modify the shape of their vocal tract by adjusting its movable parts, known as articulators [1,2], which alters the resonances of the vocal tract. When a resonance is brought close to a harmonic of the voice source, the amplitude of that harmonic, and hence of the overall sound produced, is increased, an important consideration for opera singers who must regularly perform without amplification to audiences of hundreds or even thousands. Female singers are able to sing at fundamental frequencies in excess of 1 kHz, which makes analysis of vocal tract resonances from the acoustic spectrum difficult due to the wide spacing of the harmonics. Neither spectral analysis nor linear prediction (popular in speech analysis) is reliable for detecting resonances at fundamental frequencies above approximately 350 Hz [4].

To overcome these issues, methods for directly measuring the vocal tract resonances have been developed. These methods include excitation of the vocal tract by using an external vibrator [2, 5] or by injecting a noise source, either broadband noise [1, 5-9] or swept sine [10, 11], at the lips and rerecording this to produce a transfer function of the vocal tract. Analysis of vocal fry has also been used to determine the vocal tract resonances [12]. These techniques overcome the problem of widely spaced harmonics of the voice source in the high soprano range.

To investigate the methods used by singers to produce these vocal tract resonances, the shape of the vocal tract can be measured directly using magnetic resonance imaging (MRI). This approach is particularly useful in analyzing the female vocal tract as it allows information about the articulators to be gathered over the singer's entire voice range. A number of studies have used MRI to investigate speech and singing [11, 13-15]; however, there is very little research using three-dimensional (3D) imaging to specifically investigate resonance tuning in soprano voices. Although research into the effects of various articulators on *speech* has been ongoing for over 40 years, for example [16-18], it cannot be assumed that the same articulatory techniques are used in singing, especially considering the specific challenges faced by sopranos when singing at very high fundamental frequencies, which lie well beyond the range of normal speech.

Two-dimensional (2D) MRI allows images to be captured in real time, which is closer to normal voice production; however, images from 3D MRI, although static, allow data in the transverse as well as midsagittal plane to be collected over a range of pitches. This 3D data can be used to generate more accurate cross-sectional area functions [19], and allows information such as the width of the pharynx and other adjustable parts of the vocal tract (e.g. tongue) and the volume of the vocal tract to be examined over a singer's entire pitch range. With the wealth of information available from MRI, there is a danger of becoming inundated with too many variables, which could lead to any trends in the data becoming buried in variance. It is crucial therefore, to determine the most useful and meaningful measurements in reference to resonance tuning. Combining previous work concerning vocal tract characteristics related to fundamental frequency, for example, tongue height and jaw opening [20], with newly available measurement techniques could identify useful avenues for exploring this type of data.

Previous studies on the singing voice using MRI include Echternach et al's study [21], which used real-time 2D MRI to investigate registers in the female singing voice, considering factors including lip opening, jaw opening, tongue height, jaw protrusion, oropharynx width, and uvula elevation. In a subsequent study, Echternach et al [22] also used a combination of real-time 2D and static 3D MRI to investigate 3D factors including the tongue shape, the size of the piriform sinuses, and the lip and jaw opening at very high fundamental frequencies. Bresch and Narayanan [23] used real-time 2D MRI to investigate resonance tuning in five sopranos, and although subjects generally showed a more open-mouth shape with increasing fundamental frequency, it was suggested that sopranos might not all employ the same generalizable strategies for resonance tuning as previously thought. Studies on resonance tuning, but not involving MRI, have also considered lip opening and lip spreading [24], whereas other studies on soprano singing have also considered larynx height [25, 26].

There is a precedent in this research area for studies with limited subject numbers; for example, in Sundberg [2] and Carlsson and Sundberg's [3] early work identifying resonance tuning, only one soprano was considered. Similarly, Echternach et al. used MRI to study the vocal tract of a single soprano singing at very high frequencies [22] and register changes in one tenor and one baritone [27]. Miller et al. [12] used a bass-baritone singer to compare methods of locating formant frequencies, and Delvaux and Howard [11] used one female and two male singers to investigate the impact of the piriform fossae on the singing voice. Similarly, in studies on speech, Sulter et al. [28] used a single male subject to compare predicted resonances with measured values, and Clément et al. [29] used one male speaker to compare vocal tract resonances obtained from recorded speech with those calculated from an area function of the vocal tract acquired using MRI.

Following on from the practices established in previous studies involving soprano singing and MRI methods, the principal aim of the present study was to design and test a novel protocol to investigate the vocal tract characteristics that result in resonance tuning (rather than to determine exactly how resonance tuning is affected by articulatory parameters). Measurements were taken from a single subject to test the practicalities and usefulness of this protocol, which combines direct measurements of vocal tract resonances with 3D MRI imaging, drawing on parameters identified in previous studies [21, 23-25]. Another contribution of the present study is that by using 3D MRI, it obtains transverse measurements in addition to the midsagittal plane information reported in previous studies and provides methods for quantitative analysis using these data.

METHOD

In this study, one professional singer was asked to phonate vowel sounds across her entire vocal range, both in an MRI machine and in an anechoic chamber, where the MRI tasks were repeated and measurements of her vocal tract resonances were taken.

SUBJECT

The singer used in the present study was a mezzo-soprano International Opera Principal, scoring 2.1 on the Bunch-Chapman scale [30]. She was 57 years old, and indicated a normal singing range of approximately two and a half octaves, from G3 to D6.

RESONANCE DETECTION

A method initially developed by Epps et al. [6] and used by others including Henrich et al. [1], Dowd et al. [7], Joliveau et al. [8] and Garnier et al. [9] was used to measure the resonances of the vocal tract. This consisted of exciting the vocal tract at the mouth with a synthesized broadband signal while also recording the response with a lavalier microphone placed at the subject's mouth (see Figure 1). The experimental setup for the present study is identical to that presented in Vos et al. [31] (see Figure 1).

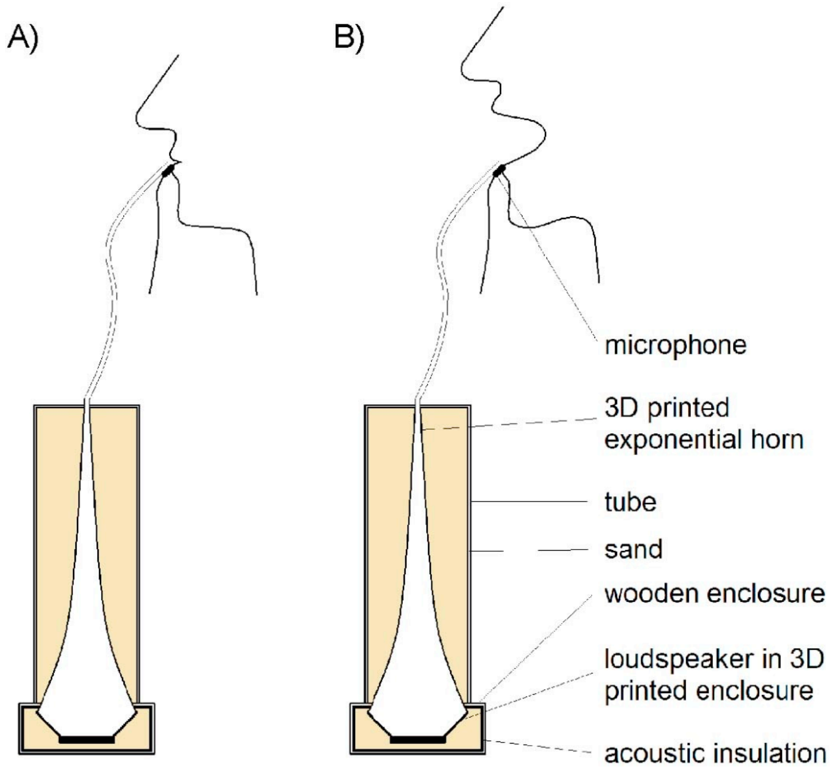


Figure 1: The equipment used to simultaneously play and record a signal at the subject's mouth using a 3D-printed impedance-matching horn and a microphone (A - calibration, B - recording). The impedance-matching horn is encased in a wooden enclosure filled with sand. The flexible tubing allows the subjects to position the acoustic source and microphone on their bottom lip.

The equipment used to simultaneously play and record a signal at the subject's mouth using a 3D-printed impedance-matching horn and a microphone (A - calibration, B - recording). The impedance-matching horn is encased in a wooden enclosure filled with sand. The flexible tubing allows the subjects to position the acoustic source and microphone on their bottom lip.

The device was held by the subjects, touching their bottom lip. The excitation signal used consisted of harmonics spaced 5.38 Hz apart, from 250 to 3500 Hz, with phases adjusted to improve the signal-to-noise ratio [32].

First, a calibration procedure was carried out. This procedure involved measuring the pressure response at the mouth with the subject's mouth closed (P_{closed}), and adjusting the amplitudes of the frequency components to make the signal strength of the microphone at the subject's mouth independent of frequency. The amplitude of each frequency component in the input signal was adjusted so that when the signal was recorded with the subject's mouth closed, the amplitudes of each frequency component became equal.

This calibrated signal was then used as the excitation signal for the measurements taken while the subject sang the required note (P_{open}). Because the source approximates an ideal current source [6], the ratio of (P_{open}/P_{closed}) therefore measures the ratio of the impedance of the vocal tract to that of the radiation field [1]. The spectrum of the signal recorded at the subject's mouth therefore shows the harmonics of the voice source superimposed on an approximate transfer function of the vocal tract. An advantage of this method is that, in addition to being reliable for measuring vocal tract resonances, it also allows the subject to sing normally while the measurement is taken. The average amplitude of the excitation signal was 84.75 dB, which introduced sufficient acoustic energy to get a reliable resonance measurement while still being low enough to allow the subjects to hear themselves, to cause minimal interference.

EXPERIMENT

Before the experiment, the singer was asked to answer a questionnaire concerning details of their singing experience and training. The singer was also asked to describe the technique she employed to sing vowels at the top of her range, and if she were aware of any differences in technique or performance when singing in a supine position.

The subject was also asked to complete consent forms for all parts of the experiment, and a safety checklist for the MRI scan. Prior ethical approval was gained from the Physical Sciences Ethics Committee at the University of York.

Vowel	/a:/	/u:/	/i:/
C4	✓	✓	✓
E4	✓	✓	✓
G4	✓	✓	✓
C5	✓	✓	✓
E5	✓	✓	✓
G5	✓	✓	✓
A#5	✓	✓	✓
C6	✓ (poor quality)	✓	✓

Table 1: The fundamental frequencies for each vowel investigated.

The first part of the procedure involved taking MRI scans of the subject. Once positioned in the MRI machine, the vocal tract was first scanned as the subject maintained a neutral vocal tract shape, described as "a relaxed neutral shape, with your mouth slightly open, breathing normally." The subject was then asked to phonate notes on three different vowels at seven pitches (see table 1). Before each scan, a recording of the target note played on a piano was played over the intercom. The scan duration was 16 seconds per note, which required the singer to maintain the shape of her vocal tract during this time, with a target phonation time of 16 seconds. The MRI machine used was a GE 3-Tesla HDx Excite MRI scanner (GE, Boston, MA), based at York Neuroimaging Centre. After the MRI scan, the subject was encouraged to take a break with food and drink as required, before proceeding to the second part.

The highest fundamental frequency investigated for the /ɑ/ vowel (C6) was discarded due to poor quality. The G4 measurement for the same vowel was initially thought to be of poor quality and was repeated, but was later found to be adequate and was included in the study.

The second part of the procedure was carried out in a fully anechoic chamber, to obtain clean audio recordings of the same sounds over a greater range of pitches, without the presence of MRI noise. The singer was asked to lie supine on a foam-covered board and wear headphones playing recorded MRI noise while singing (not audible on the recording), to simulate the conditions in the MRI machine. The singer was first asked to sing individual notes, each on one breath, in an ascending chromatic sequence (12 notes per octave) from C4 to the top of their range, singing into the wide-band vocal tract measuring device (see

Resonance Detection). The singer was required to hold each note for approximately 6 seconds, and each note was given on an electric piano before it was sung. This was then repeated on each vowel sound (/a:/, /u:/, and /i:/). The subjects were asked to sing in their "normal performance voice," keeping their mouth shape constant for the duration of each measurement and at a medium level. The subjects were reminded if necessary during the tasks. Notes were only repeated if the measurement was insufficient or if the subject failed to maintain the note until the end of the measurement.

ANALYSIS

The images obtained by MRI were imported into *ITK-SNAP* [33], and the "annotation" tool was used to directly measure the dimensions of the vocal tract in the midsagittal plane. After Echternach et al. [21] the parameters measured were the (a) lip opening, (b) jaw opening, (c) height of the tongue dorsum, (d) jaw protrusion, (e) oropharynx width, and (f) uvula elevation; in addition to these, the (g) oropharynx breadth (perpendicular to the oropharynx width), (h) larynx height, (i) lip spreading, and (j) vocal tract length (the length of the midline of the vocal tract calculated with the area function—see Generation of 3D Area Function) were measured. The larynx height was measured by taking the distance of the larynx to a fixed point (the collarbone) for all sung notes and the "neutral" position, then subtracting the distance for the neutral position. The midsagittal measurements are shown in Figure 2.

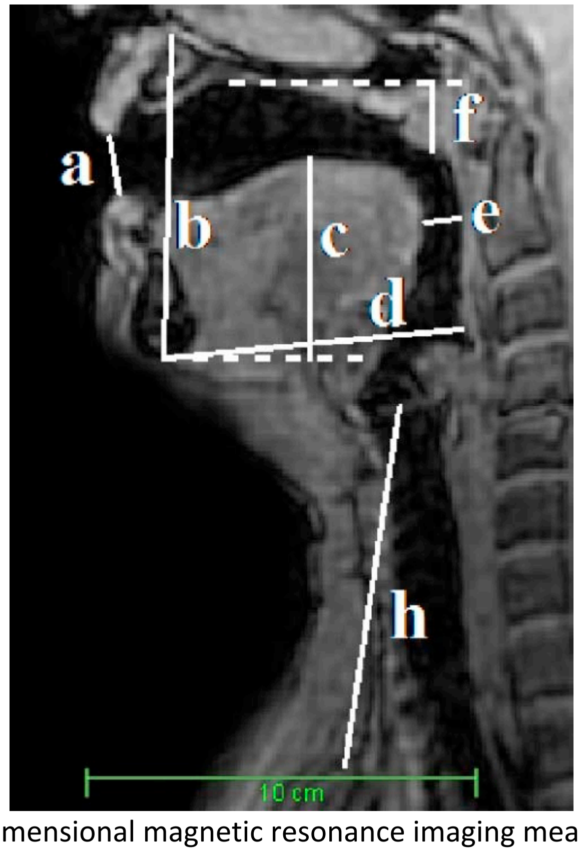


Figure 2: Two-dimensional magnetic resonance imaging measurements: (a) lip opening, (b) jaw opening, (c) height of the tongue dorsum, (d) jaw protrusion, (e) oropharynx width, (f) uvula elevation, and (h) larynx height. (g) Oropharynx breadth, (i) lip spreading, and (j) vocal tract length are not shown.

Using *ITK-SNAP*, the airway was segmented to produce a 3D vocal tract volume, and the radiation dome was removed. This segmentation was then imported into *ParaView* [34] and exported as a list of 3D points on the surface of the vocal tract, as well as connectivity data for the points to be loaded into *MATLAG* [35] (The MathWorks Inc., Natick, Massachusetts) for analysis. The x-direction was defined as transverse (left-right), the y-direction as anterior-posterior (front-back), and the z-direction as superior-inferior (up-down). All measurements were taken in millimeters.

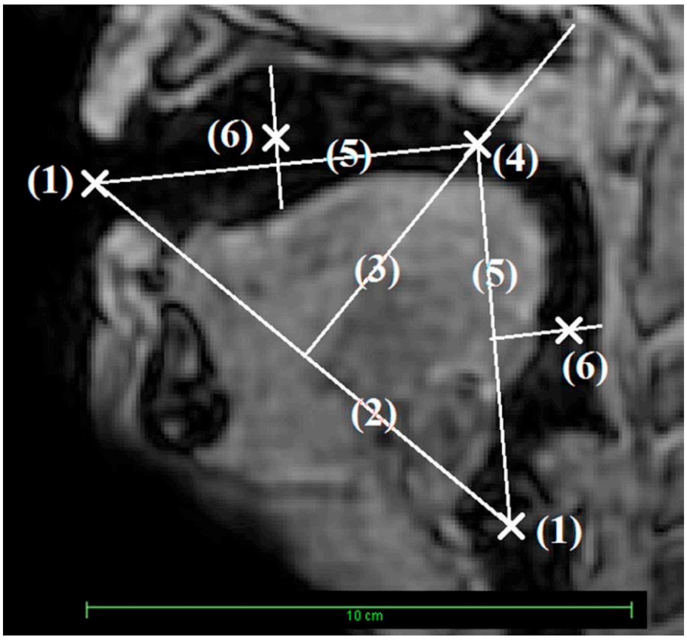
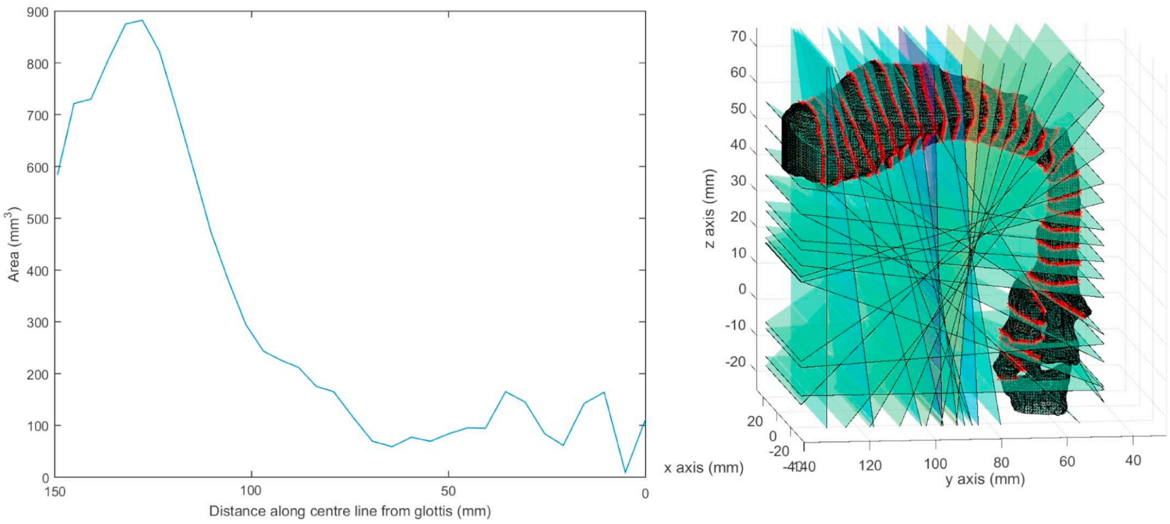


Figure 3: Illustration of the algorithm to determine the slicing of the vocal tract.

The start (glottis) and end (mouth) of the vocal tract were manually defined by the researcher, labeled as (1) in Figure 3, and then following an algorithm originally developed to analyze upper airway geometry and volume with regard to sleep disorders [36] and adapted to generate a 2D area function from a midsagittal slice [37], the area function was calculated using an iterative bisection algorithm: first, the line joining the start and the end of the vocal tract was calculated (2), and then a plane was defined at the midpoint of this line, normal to it (3). The intersection of this plane with the vocal tract was found, and its area and center were then calculated (4), and the center was stored as a point on the midline of the vocal tract. This process was then repeated between the start of the vocal tract and the midpoint, and between the midpoint and the end, "slicing" the vocal tract into quarters (5), and again finding the areas and midpoints of these intersections (6).



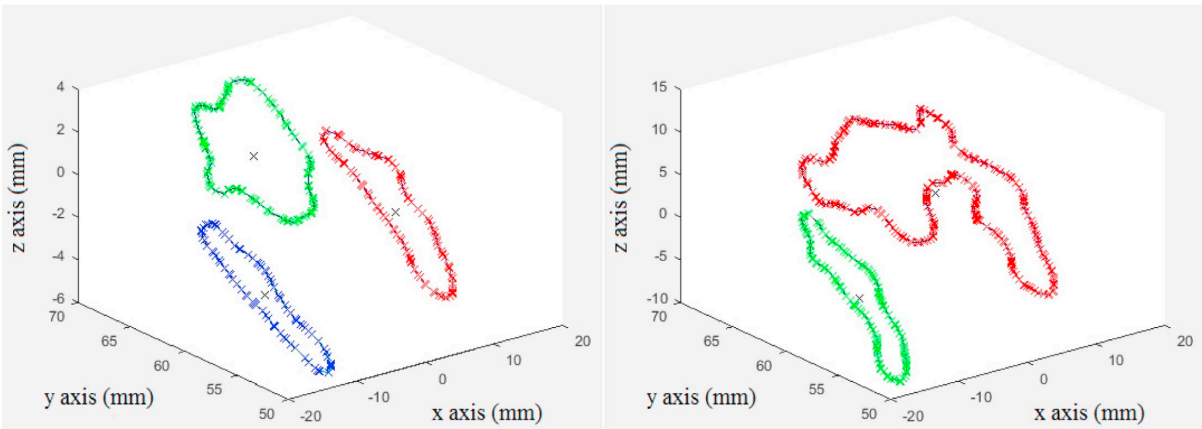
(A) Area function

(B) 33 Slicing planes

Figure 4: An example of an area function generated (A) and the planes used to generate it (B).

This slicing could then be repeated (slicing into 8ths, 16ths, etc) to produce a vocal tract cut into 2^n parts. The areas of the start and end points were also included, with the first slicing plane defined as horizontal ($x - y$) and the last as vertical ($x - z$). This yielded an area function of $2^n + 1$ slices; in the present study, n was chosen to be 5, giving 33 slices in total. This was found to provide a sufficient level of detail for analysis while not taking an excessively long time to calculate. An example of the 3D vocal tract mesh, with the planes used to slice it, is shown in Figure 4A, and the area function generated by this is shown in Figure 4B .

A number of restrictions were implemented in this procedure to make the process more reliable; first, the x-component of the center of each area slice was restricted to the midpoint of the previous and following x-components. In addition to this, the slicing plane was forced to face forward (x-component of the normal made zero) to reduce the likelihood of areas overlapping with the previous or following ones.



(A) Slice through the 4th plane from the glottis.

(B) Slice through the 6th plane from the glottis.

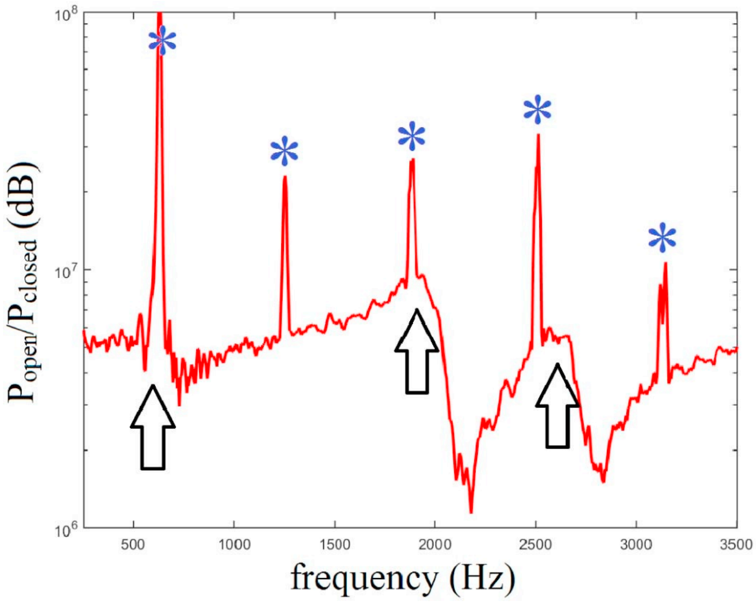
Figure 5: An example of two slices through the vocal tract used to generate the area function: the 4th (A) and the 6th (B) slices.

Some difficulty was encountered in analysis due to the piriform fossae, as in some cases, the intersection of the vocal tract with the "slicing plane" produced more than one area. If there was more than one separate area identified, then the most central one was chosen, and its area was calculated. Due to the slight asymmetry of the piriform fossae, however, this meant that occasionally one (or part of one) of them was included in the area (as it was not quite separate from the main area of the vocal tract), whereas the other one was discarded. This led to some error in the measurements of the cross-sectional area, in the region around 1–2 cm from the glottis. An example of this is shown in Figure 5. With the same measurement as Figure 4, the fourth plane from the glottis slices through three separate areas (Figure 5A); however, the sixth plane (Figure 5B) only identifies two areas.

Although the resonances of the vocal tract could be calculated directly from the area functions generated from MRI images, this would not take into account effects such as the radiation impedance at the subject's mouth, or the wall compliance within the vocal tract. Since the resonance measurements made in this experiment (using broadband noise excitation) measure the resonances directly, they can be assumed to be taking these effects into account.

RESONANCE TUNING MEASUREMENTS

The lead author manually determined the frequencies of the vocal tract resonances, from the broad peaks in the plots of P_{open}/P_{closed} against frequency. An example plot of P_{open}/P_{closed} against frequency is shown in Figure 6. As in previous studies [1, 9, 31], these measurements were then cross-checked by another researcher. In some cases, it was not possible to accurately identify the vocal tract resonances, especially for closed vowels or when the subject did not remain completely still while singing[1] and these measurements were omitted from the results.



[1] In some cases, this could be identified by observing the subject; however, movement of the subject also produced a characteristic error in the measurement, which allowed this to be detected.

Figure 6: A plot of P_{open}/P_{closed} against frequency for an /u:/ vowel. The first five harmonics are marked with asterisks, and the first three resonances are marked with arrows.

RESULTS

All the MRI measurements, fundamental frequencies, and the measurements of the first and second resonances of the vocal tract (R_1 and R_2, respectively) while the singer was singing in a supine position in the anechoic chamber were imported into MATLAB [35] for statistical analysis. The linear correlations between all the MRI measurements and R_1 and R_2 were calculated, and a correlation matrix was generated (see Figure 7). The results that were not significant at the 5% level were omitted from the matrix. Significant positive correlations are represented as striped, whereas significant negative correlations are represented as dark gray. The raw data (MRI measurements) associated with this experiment are available online [38].

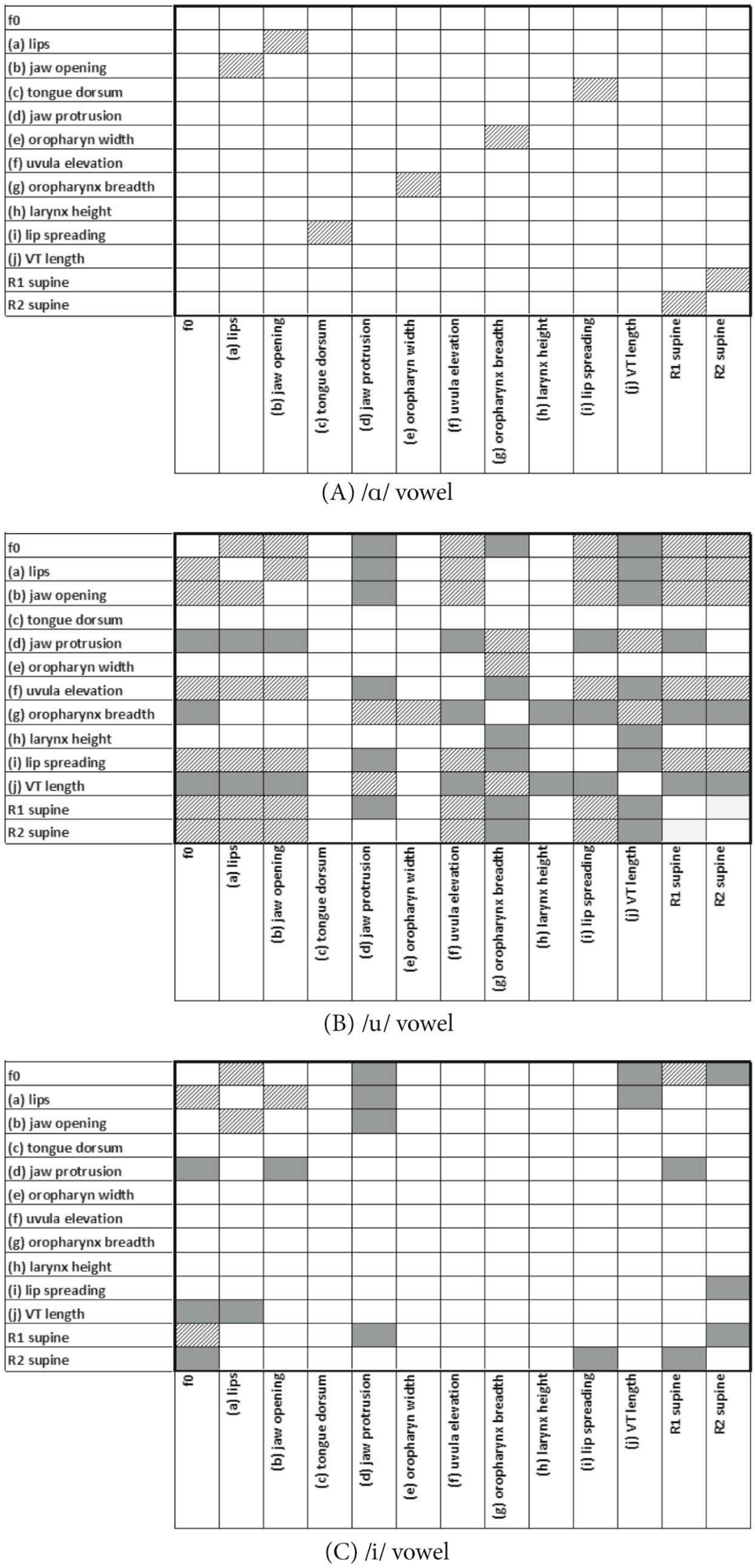


Figure 7: Correlation matrices for all variables, with nonsignificant results (at 5% level) removed. Positive correlations are represented as striped, whereas negative correlations are represented as dark gray. (A) /a:/ vowel, (B) /u:/ vowel, (C) /i:/ vowel.

The correlation matrix showing the correlation between all the MRI measurements and R_1 and R_2 can be seen in Figure 7. The only correlation consistent across all three vowels is a significant positive correlation between (a) the lip opening and (b) the jaw opening, which is expected, as both of these measures describe the degree of openness of the singer's mouth.

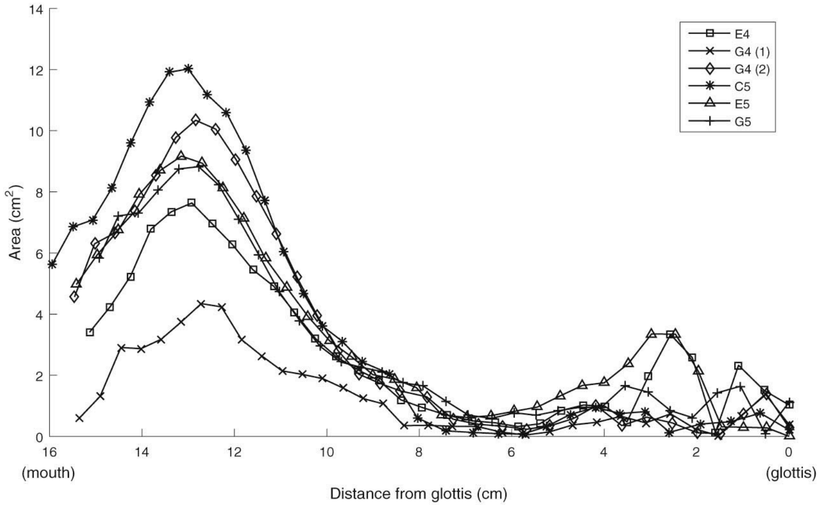
For the /a:/ vowel, there were no significant correlations between the fundamental frequency and any of the other measurements. The first and second resonances only showed a significant correlation with each other and not with any of the other variables. The only other significant correlations were between the (c) tongue dorsum and (i) lip spreading, and between (e) oropharynx width and (g) oropharynx breadth.

For the /u:/ vowel, a great deal more correlation between the variables was seen than for either of the other two vowels (45/78 significant correlations, compared to 4/78 for /α/, and 12/78 for /i:/). The fundamental frequency showed a positive correlation with the (a) lips, (b) jaw opening, (f) uvula elevation, (i) lip spreading, and R_1 and R_2. A negative correlation was seen between fundamental frequency and (d) jaw protrusion, (g) oropharynx breadth, and (j) vocal tract length.

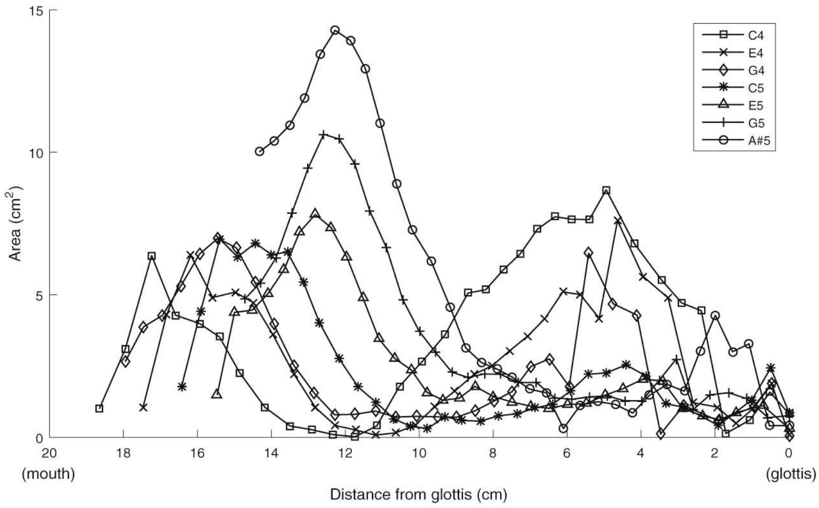
Although the tongue position is generally accepted to affect the position of R_2 [20], for this subject there was no linear correlation between the tongue dorsum and any other variable. R_1 and R_2 both showed a correlation with several other variables, which were the same except for the addition of (d) jaw protrusion for R_1. Not surprisingly, both resonance measurements showed a negative correlation with the (j) vocal tract length, supporting the acoustic principle that shortening a pipe raises the frequencies of its resonances.
The /i:/ vowel showed less correlation overall than the /u:/ vowel, but a little more than the /a:/ vowel. Contradictory to the results for the /u:/ vowel, the R_1 and R_2 measurements showed completely different correlations, although they both correlated with fundamental frequency, which was positive for R_1 and negative for R_2.

R_1 and R_2 showed a correlation with several variables for both /u:/ and /i:/ vowels, but not always in the same way. For example, for the /u/ vowel, R_2 showed a positive correlation with lip spreading, whereas for the /i:/ vowel, R_2 had a negative correlation with this variable.

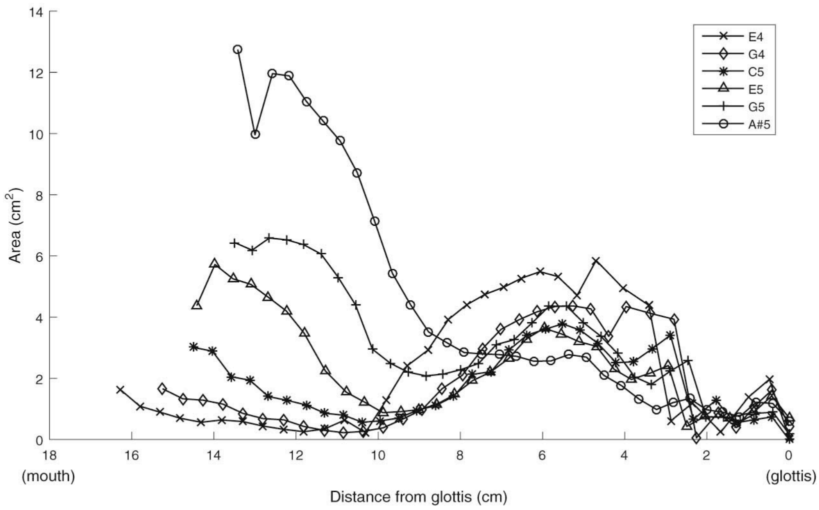
AREA FUNCTIONS
The area functions were grouped by vowel and then plotted on the same axes (see Figure 8) to allow patterns in the data to be seen.



(A) /ɑ/ vowel



(B) /u/ vowel



(C) /i/ vowel

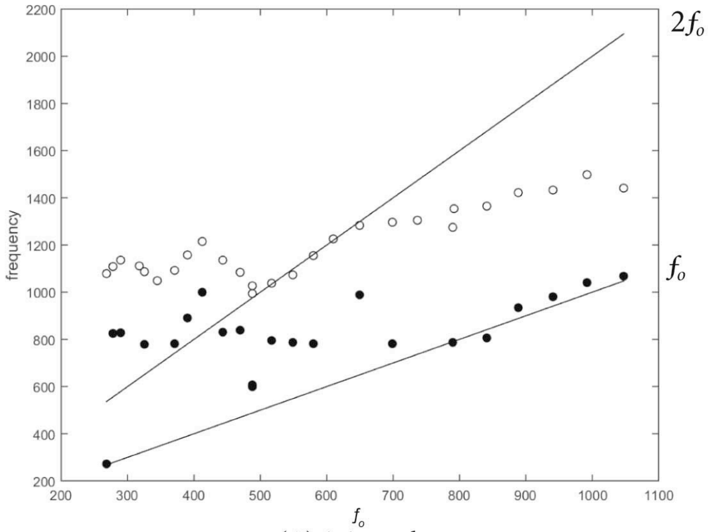
Figure 8: Area functions for all pitches, for (A) the /a:/ vowel, (B) the /u:/ vowel, and (C) the /i:/ vowel.

The area function for the /a:/ vowel was characterized by an approximately bell-shaped vocal tract: narrowing to approximately 1 cm^2 around 6–7 cm from the glottis (around the back of the tongue), then opening out around 13 cm, before narrowing at the mouth. Although the extent of the mouth opening varied for different fundamental frequencies (between 4 and 12 cm^2), there did not appear to be any relationship between fundamental frequency and mouth opening.

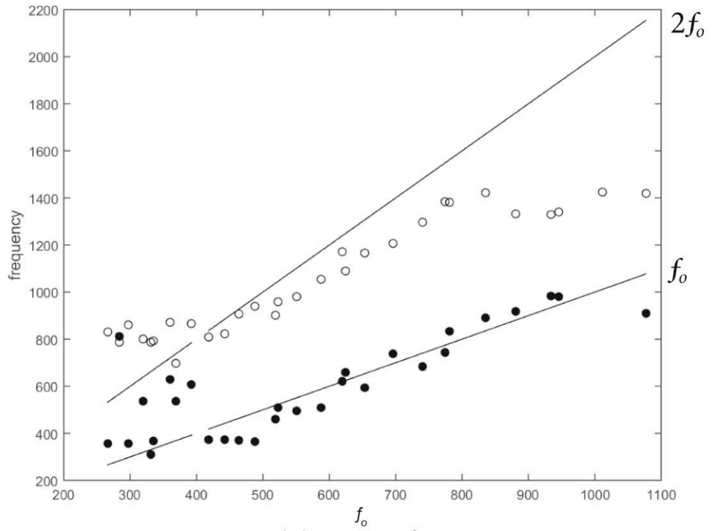
Interestingly, for the /u:/ vowel, the lower pitches showed a large space of about 8 cm^2 around the pharynx (approx. 5 cm from the glottis), which then decreased to a very small cross-sectional area around 12 cm from the glottis, and then opened up a little before a final restriction at the mouth. For the higher fundamental frequencies, the shape was very similar to the /a:/ vowel, with a narrowing around 6 cm, then a large opening up to approximately 14 cm^2, before a slightly smaller mouth area. At certain points along the vocal tract, there appeared to be a relationship between the cross-sectional area and the fundamental frequency. For example, at around 5 cm from the glottis, the lowest fundamental frequency had the highest area, and the highest fundamental frequency had the lowest area. The opposite effect was seen at 13 cm from the glottis, where the highest fundamental frequency had the lowest area and vice versa. A noticeable shortening of the vocal tract was also seen with increasing fundamental frequency, possibly due to the corners of the mouth being pulled back, changing its effective length.

The same patterns between the cross-sectional area and fundamental frequency were also seen for the /i:/ vowel: at the mouth, where the lowest fundamental frequencies had the lowest cross-sectional areas; 4–6 cm from the glottis (pharynx), where the lowest fundamental frequencies had the highest areas (approximately 6 cm^2); and with the shortening of the vocal tract with increasing fundamental frequency.

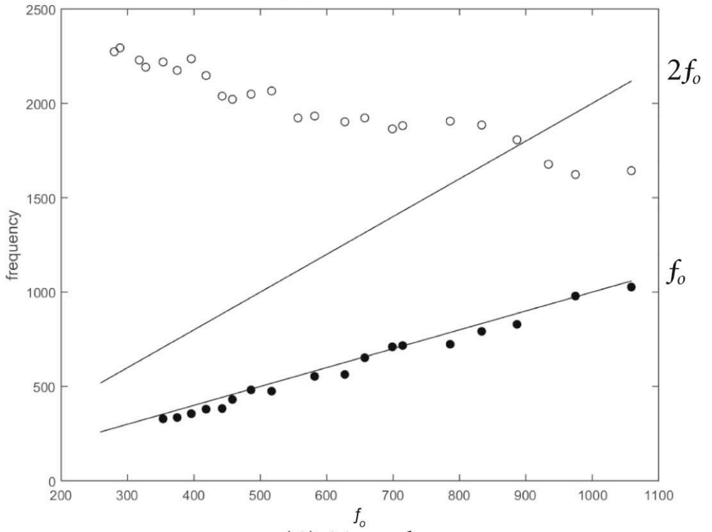
RESONANCE MEASUREMENTS



(A) /ɑ/ vowel



(B) /u/ vowel



(C) /i/ vowel

Figure 9: Measurements of R_1 (filled circles) and R_2 (open circles) against fundamental frequency for all pitches, sung in the anechoic chamber, in the supine position, for the (A) /a:/ vowel, (B) /u:/ vowel, and (C) /i:/ vowel. The solid lines show the first and second harmonics.

A plot of the measured resonances for each vowel is shown in Figure 9, with R_1 represented as filled circles and R_2 as open circles. These results are summarized in Figure 10, which shows the range and extent of both $R_1 : f_o$ and $R_2 : 2f_o$ tuning over the range of fundamental frequencies investigated, to within 70 Hz (in gray, after Vos et al[31]), and "tighter" resonance tuning, to within 25 Hz (in black, after Henrich et al[1] and Garnier et al[9]).

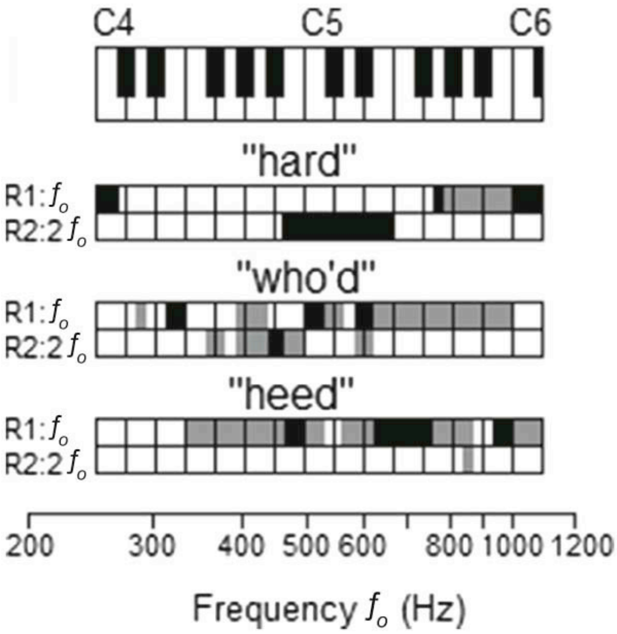


Figure 10: Shaded boxes show resonance tuning for each vowel (/a:/, /i:/ and /u:/), to within 70 Hz (gray) or 25 Hz (black). The top line for each vowel shows $R_1 : f_o$ tuning, bottom line $R_2 : 2f_o$.

The measurements of R_1 appear very scattered for the /a:/ vowel, especially toward the bottom of the singer's range, with $R_1 : f_o$ tuning only seen in approximately the top third of fundamental frequencies investigated. $R_2 : 2f_o$ tuning was observed only briefly, just below the middle of the range of fundamental frequencies investigated.

For the /u:/ vowel, $R_1 : f_o$ tuning was seen over nearly the entire range of fundamental frequencies investigated, albeit only "loosely" (to within 70 Hz of f_o). Over some of this range, $R_2 : 2f_o$ tuning was observed in conjunction with $R_1 : f_o$ tuning, ceasing at approximately D#5 (622 Hz).

The resonance measurements for the /i:/ vowel appear to follow very clear patterns; $R_1 : f_o$ tuning was seen over the entire fundamental frequency range, and R_2 descended as the fundamental frequency increased. The resonance measurements for this vowel showed the least scattering of all three vowels investigated.

Analysis	/a:/	/u:/	/i:/

2d MRI measurements.	Little correlation seen (4/78 pairings), no correlations with R_1 or R_2.	Great deal of correlation (45/78 pairings), correlations with R_1 and R_2 similar.	Some correlation seen (12/78 pairings), different correlations with R_1 and R_2.
Area functions.	Bell-shaped curve, no pattern with fundamental frequency.	Higher fundamental frequencies show larger mouth opening, smaller pharynx space, and shortening of the vocal tract.	
Resonance measurements.	$R_1 : f_o$ tuning seen at the top of the range, $R_2 : 2f_o$tuning in the middle of the range.	$R_1 : f_o$ across most of the range, small amount of $R_2 : 2f_o$ tuning in the middle of the range.	$R_1 : f_o$ tuning across a wide range, no $R_2 : 2f_o$ tuning, but R_2 showed a strong negative correlation with f_o.

Table 2: Summary of the 2D MRI Measurements, Area Functions, and Resonance Measurements

A summary of the 2D measurements and their correlations, area functions, and resonance measurements for each vowel is shown in Table 2.

DISCUSSION

Different levels of correlation between variables (none for the /a:/ vowel, many for the /u:/ vowel, and f_o, (d) jaw protrusion, and (i:) lip spreading for the /i:/ vowel) with R_1 and R_2 indicate not only that this singer used different techniques to produce different vowels but also that the effect of changing one variable would depend on the vowel sung. For instance, for the /u:/ vowel, R_2 showed a positive correlation with lip spreading; however, for the /i:/ vowel, R_2 had a negative correlation with this variable.

Considering the other correlations for the /a:/ vowel, a correlation between the width and breadth of the oropharynx may imply a causal relationship. However, the correlation between lip spreading and the tongue dorsum seems less likely to be due to a causal relationship between these two variables, and these factors may both be dependent on a third factor.

The only variables that did not show a correlation with resonances for any vowels were the (c) tongue dorsum, (e) oropharynx width, and (h) larynx height, which could suggest that these variables are not of interest when considering resonance tuning in one of the three vowels investigated; however, data from additional subjects would be required to verify this.

It has been noted [39] that singers generally find the /a:/ vowel the easiest and most natural to sing, as it does not require the extreme vocal tract adaptations required for the /i:/ vowel (which is generally found difficult to sing, especially at high pitches), and this may be a reason for the lack of correlation between variables seen for the /a:/ vowel.

The singer commented before the experiment that she used “the same technique for all vowels but /i:/ was the hardest,” whereas after the experiment, she said that she “found /a:/ the hardest to sing high up, whereas it would normally be /i;/.” This theory is supported by the area functions for the /a:/ vowel; there may be no clear dependence of mouth opening on pitch seen for the /a:/ vowel because the singer does not need to make a special effort to produce this vowel, unlike for the more difficult vowels, /u:/ and /i:/, which showed a clear pattern.

The pattern in resonance measurements for the /i:/ vowel could also be linked to difficulty. Even though R_2 is not strictly tuned, it shows a clear relationship with fundamental frequency; the resonance measurements for the /i:/ vowel show the least variation out of all three vowels. Singers typically find an intelligible /i:/ at a high fundamental frequency very difficult to sing as it requires a very closed mouth shape, which limits the amplitude of the sound produced. Producing a loud but intelligible /i:/ must therefore be a trade-off between these two perceptual attributes. This could mean that the stricter acoustic requirements for producing an /i:/ vowel mean that there is less room for variation in technique, so unlike the /a:/ vowel, even when in an unusual situation, the vowel is still produced in a very consistent fashion.

Jaw and Tongue

$R_1 : f_o$ tuning was seen over a wide range of fundamental frequencies for the /u:/ and /i:/ vowels, which may be due to the larger mouth area observed with increasing pitch, supporting the theory that jaw opening lowers R_1 [20]. Surprisingly, however, the correlation matrices did not show a correlation between fundamental frequency and jaw opening for the /i:/ vowel, which is suggested from the area functions. This may be due to insufficient data to produce statistical significance, or a nonlinear relationship (see Nonlinear Effects).

Before the experiment, the singer said she was unaware of changing her technique when lying down, and described her technique for singing high notes as "relaxed jaw and lifted palate. Firm support from pelvic area (tilt pelvis forward and unlock knees)." This suggests that there should be a correlation between fundamental frequency and jaw opening; however, this was seen only for the /u:/ vowel and not for the other two vowels investigated. When asked if she was aware that she made changes to the shape of her vocal tract when singing high notes, the singer said that she also "brought her tongue forward and down as she sung higher." However, no correlation between the fundamental frequency and the tongue dorsum was seen for any of the vowels.

Although the tongue position is generally accepted to affect the position of R_2 [20], for this subject for the /u:/ and /i:/ vowels, there was no significant correlation (at the 5% level) between the tongue dorsum and any other variable. This finding suggests that this particular singer did not make use of this technique during this experiment; however, it cannot be known whether this reflects her usual technique, only her performance during this investigation. After the experiment, the singer commented that in the MRI scanner, she was very aware of her jaw being "tense" and that she felt her tongue "was further back than normal." It is possible that the effects of lying down in the scanner due to the altered effects of gravity [40, 41], and the restrictive position, impeded the singer's normal vocal tract adjustments, possibly stopping her from using her tongue to tune her second resonance where she would normally have employed this technique.

The area functions of the highest fundamental frequencies are similar for all three vowels, which agree with the idea that singers make use of similar vocal tract positions across vowels at the top of their range [39].

The area functions for these two vowels also suggest a relationship between space around the oropharynx and fundamental frequency; however, the only correlation between oropharynx measurements and fundamental frequency occurs for the /u:/ vowel, when it correlated with oropharynx breadth and also R_2.

Results for /u:/ and /i:/ vowels agree with the findings by Bresch and Narayanan [23], who observed (from midsagittal images) that “the front cavity opens more widely as the singer goes to higher fundamental frequencies.”

However, Bresch and Narayanan observed this behavior across all five vowels investigated (/a:/, /e/, /i:/, /o/, and /u:/), which may suggest that the singer in the current study is exhibiting atypical behavior; either she has an unusual technique or she was unable to use her normal techniques due to the restrictions of the MRI machine.

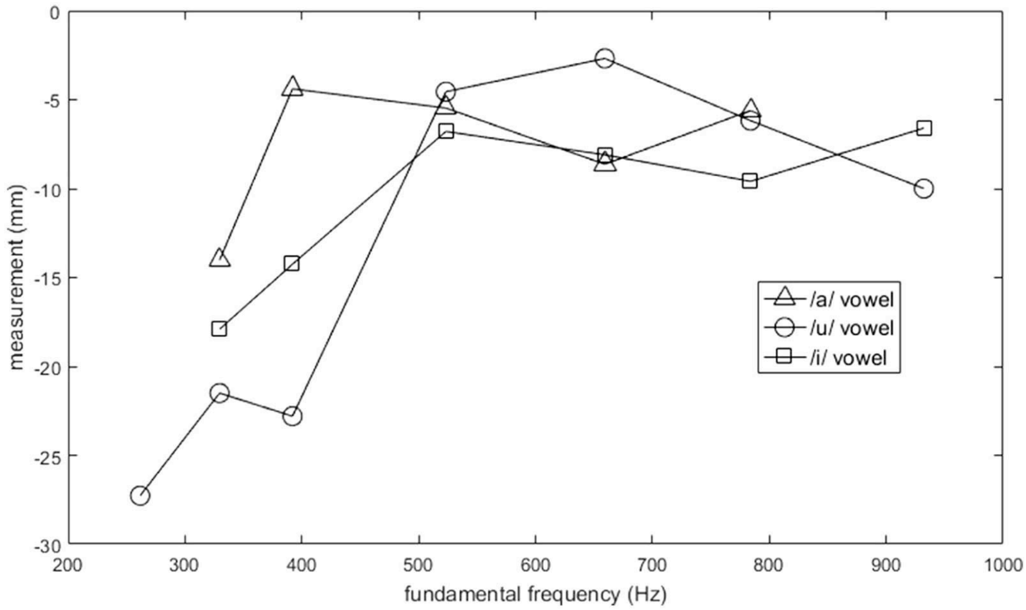


Figure 11: Larynx heights against fundamental frequency for each vowel.

The present study has only examined *linear* correlations between variables, and *nonlinear* relationships may exist. Based on the acoustic properties of standing waves in tubes, it was expected that a simple linear relationship would be seen between the vocal tract resonances and factors that cause shortening or lengthening, or constriction or expansion of the vocal tract, such as the jaw opening, tongue height, or larynx height. The larynx height, however, followed a similar pattern for all three vowels (see Figure 11), increasing at first with fundamental frequency but then remaining approximately constant across the top half of the fundamental frequency range investigated. This does not agree with previous studies [25, 26], which found that for sopranos, the larynx generally rose with fundamental frequency in the upper or top part of the range, although differences were seen between individual singers.

It may be necessary to consider the interplay between variables more carefully. For example, the vocal tract length depends on both the larynx height, as a higher larynx shortens the vocal tract, and the lip spreading, as when the corners of the mouth are pulled back; this also effectively shortens the vocal tract.

There are a number of limitations to be considered in the present study, which mostly arise from the conditions necessary for MRI; first, the singer must be supine, and was strapped to a board and unable to move for the duration of each measurement, and was required to sustain

each note for an unnaturally long time. All of these factors are unnatural for the singer and could possibly have an unknown effect on the measurements obtained.
Only a single subject was used in the present study, which makes it difficult to separate individual habits from general trends. However, choosing a highly trained professional singer alleviates some of these concerns; professional singers are likely to be very reliable in their technique, so repeat measurements may not be necessary. It should also be remembered that opera involves acting as well as singing, so professional opera singers are also used to singing in unusual situations, including a supine position, so although this is not standard practice, it may not be entirely unusual.

Although the present study used a single subject, to test a suitable protocol for identifying resonance tuning techniques using 3D MRI measurements, the results from this one subject are interesting in themselves, as they provide a detailed insight into the movement of the articulators of a very high-quality singer (2.1 on the Bunch-Chapman taxonomy [30]). It is not uncommon for studies on the singing voice to use very few subjects [2, 21, 27, 28]; however, future work will expand the present study to include more singers of similar voice type and experience, which will allow more robust statistical analysis and investigation into the similarities between individuals.

To more completely identify the relevant parameters for vocal tract characterization, in reference to resonance tuning, it may be necessary to introduce more variables such as the volumes of particular parts of the vocal tract (e.g. the pharynx) or more measurements in the transverse plane. It should also be noted that the resonance frequencies may be influenced by other factors not considered in this study, such as the wall compliance of the vocal tract; however, it is the large articulators such as the jaw and tongue that have the greatest effect on vocal tract shape, so these are the parameters focused on in the present study. It has therefore been assumed that factors such as wall compliance have remained approximately constant across all measurements.

CONCLUSIONS

The present study has presented a new protocol for investigating the parameters affecting resonance tuning in soprano singers. Good-quality measurements were obtained from a single subject, allowing area functions to be generated and the positions of the vocal tract articulators to be monitored and compared to the vocal tract resonances. Upon analysis, a highly complex interplay between variables was observed; there did not appear to be any clear linear relationship between the parameters extracted from MRI data and measurements of resonance tuning across all vowels.

It is not possible to generalize the results from this single singer; however, the different relationships between articulators and resonances observed for the three vowels investigated in this study support the ideas of Bresch and Narayanan [23], who suggested that sopranos might not all employ the same generalizable strategies for resonance tuning, as has been previously thought.

With the increased availability of MRI, it is crucial to understand exactly which measurements are the most relevant when considering resonance tuning in soprano singing. The detailed (3D) measurements and statistical analysis presented in this paper demonstrate a more rigorous, quantitative approach to these types of data. The present study therefore provides a baseline protocol for investigation of soprano singing using 3D MRI, with specific quantitative analyses, in consideration of the female singing voice. In particular, this has implications on future research in terms of generalizing findings across vowels, as well as informing the development of more accurate acoustic models of the singing voice.

ACKNOWLEDGEMENTS
The authors thank the Audio Engineering Society Educational Foundation for the Grant for Graduate Studies, which supported this work; the subject, for her time and effort; the staff at York Neuroimaging Centre, for their assistance with magnetic resonance imaging scanning; and John Smith and Joe Wolfe of the Acoustics Lab at the University of New South Wales, who developed and wrote the resonance measuring software.

REFERENCES

1. Henrich, N., Smith, J., and Wolfe, J. Vocal tract resonances in singing: strategies used by sopranos, altos, tenors, and baritones. *Journal of the Acoustical Society of America,* 2011; 129: 1024–1035.
2. Sundberg, J. Formant technique in a professional female singer. *Acta Acustica United Acustica*. 1975; 32: 89–96.
3. Carlsson, G. and Sundberg, J. Formant frequency tuning in singing. Journal of Voice, 1992; 6: 256–260.
4. Monsen, R.B. and Engebretson, A.M. The accuracy of formant frequency measurements: a comparison of spectrographic analysis and linear prediction. Speech, Language and Hearing Research, 1983; 26: 89–97
5. Erickson, M.L. and Alfonso, A.E.D. A comparison of two methods of formant frequency estimation for high-pitched voices. Journal of Voice, 2002; 16: 147–171
6. Epps, J., Smith, J.R., and Wolfe, J. A novel instrument to measure acoustic resonances of the vocal tract during phonation. Measurement Science Technical, 1997; 8: 1112
7. Dowd, A., Smith, J., and Wolfe, J. Real time, non-invasive measurements of vocal tract resonances: application to speech training. Acoustics Australia, Citeseer, 1996
8. Joliveau, E., Smith, J., and Wolfe, J. (number 6970)*Acoustics: Tuning of Vocal Tract Resonance by Sopranos.* Vol. 427. Nature Publishing Group, London, UK; 2004: 116
9. Garnier, M., Henrich, N., Smith, J. et al. The tuning of vocal resonances and the upper limit to the high soprano range. Proceedings of the International Symposium on Music Acoustics, Sydney & Katoomba, Australia, 2010, pp 11–16.
10. Delvaux, B. and Howard, D.M. Using an exponential sine sweep to measure the vocal tract resonances. Journal of the Acoustical Society of America, 2013; 134: 4206
11. Delvaux, B. and Howard, D.M. A new method to explore the spectral impact of the piriform fossae on the singing voice: benchmarking using MRI-based 3D-printed vocal tracts. (e102680) PLoS ONE, 2014; 9
12. Miller, D.G., Donald, G., Sulter, A.M. et al. Comparison of vocal tract formants in singing and non-periodic phonation. Journal of Voice, 1997; 11: 1–11
13. Baer, T., Gore, J.C., Gracco, L.C. et al. Analysis of vocal tract shape and dimensions using magnetic resonance imaging: vowels. Journal of the Acoustical Society of America, 1991; 90: 799–828
14. Narayanan, S., Nayak, K., Lee, S. et al. An approach to real-time magnetic resonance imaging for speech production. Journal of the Acoustical Society of America, 2004; 115: 1771–1776

15. Takemoto, H., Honda, K., Masaki, S. et al. Measurement of temporal changes in vocal tract area function from 3D cine-MRI data. Journal of the Acoustical Society of America, 2006; 119: 1037–1049
16. Fant, G. *Acoustic Theory of Speech Production: with Calculations Based on X-ray Studies of Russian Articulations.* Vol. 2. Walter de Gruyter, Berlin, Germany; 1971
17. Delattre, P. and Freeman, D.C. A dialect study of American r's by x-ray motion picture. Linguistics, 1968; 6: 29–68
18. Subtelny, J.D. and Oya, N. Cineradiographic study of sibilants. Folia Phoniatrica, 1972; 24: 30–50
19. Kim, Y.-C., Kim, J., Proctor, M. et al. Toward automatic vocal tract area function estimation from accelerated three-dimensional magnetic resonance imaging. (ISCA Workshop on Speech Production in Automatic Speech Recognition, Lyon, France, pages 2–5); 2013
20. Sundberg, J. Vocal tract resonance in singing. *J Sing*. 1988; 44: 11–31
21. Echternach, M., Sundberg, J., Arndt, S. et al. Vocal tract in female registers—a dynamic real-time MRI study. Journal of Voice, 2010; 24: 133–139
22. Echternach, M., Birkholz, P., Traser, L. et al. Articulation and vocal tract acoustics at soprano subject's high fundamental frequencies. Journal of the Acoustical Society of America, 2015; 137: 2586–2595
23. Bresch, E. and Narayanan, S. Real-time magnetic resonance imaging investigation of resonance tuning in soprano singing. Journal of the Acoustical Society of America, 2010; 128: EL335–EL341
24. Garnier, M., Henrich, N., Smith, J. et al. Vocal tract adjustments in the high soprano range. Journal of the Acoustical Society of America, 2010; 127: 3771–3780
25. Pabst, F. and Sundberg, J. Tracking multi-channel electroglottograph measurement of larynx height in singers. Logopedics Phonatrics Vocology, 1993; 18: 143–152
26. Daffern, H. Distinguishing Characteristics of Vocal Techniques in the Specialist Performance of Early Music. University of York, York, UK; 2008
27. Echternach, M., Sundberg, J., Arndt, S. et al. Vocal tract and register changes analysed by real-time MRI in male professional singers—a pilot study. Logopedics Phonatrics Vocology, 2008; 33: 67–73
28. Sulter, A.M., Miller, D.G., Wolf, R.F. et al. On the relation between the dimensions and resonance characteristics of the vocal tract: a study with MRI. Journal of Magnetic Resonance Imaging, 1992; 10: 365–373
29. Clément, P., Hans, S., Hartl, D.M. et al. Vocal tract area function for vowels using three-dimensional magnetic resonance imaging. A preliminary study. Journal of Voice, 2007; 21: 522–530
30. Bunch, M. and Chapman, J. Taxonomy of singers used as subjects in scientific research Journal of Voice, 2000; 14: 363–369
31. Vos, R.R., Daffern, H., and Howard, D.M. Resonance tuning in three girl choristers. (122-e1) Journal of Voice. 2017; 31
32. Smith, J.R. Phasing of harmonic components to optimize measured signal-to-noise ratios of transfer functions. Measurement Science Technolog,. 1995; 6: 1343
33. Yushkevich, P.A., Piven, J., Hazlett, H.C. et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage, 2006; 31: 1116–1128

34. Ahrens, J., Geveci, B., and Law, C. ParaView: an End-User Tool for Large Data Visualization. Visualization Handbook, Elsevier, The Netherlands; 2005
35. MATLAB. The MathWorks Inc. version R2016a, Natick, Massachusetts; 2016
36. Hoffman, E.A. and Gefter, W.B. Multimodality imaging of the upper airway: MRI, MR spectroscopy, and ultrafast X-ray CT. Progress in Clinical Biological Research, 1990; 345: 291
37. Story, B.H., Titze, I.R., and Hoffman, E.A. Vocal tract area functions from magnetic resonance imaging. Journal of the Acoustical Society of America,. 1996; 100: 537–554
38. Rebecca Vos. Rebecca Vos—Resources. (Available at:) http://www-users.york.ac.uk/~rrv501/
39. Miller, R. On the Art of Singing. Oxford University Press, Oxford, UK; 1996: 61
40. Shiller, D.M., Ostry, D.J., and Gribble, P.L. Effects of gravitational load on jaw movements in speech, Journal of Neuroscience, 1999; 19: 9073–9080
41. Tiede, M.K., Masaki, S., Wakumoto, M. et al. Magnetometer observation of articulation in sitting and supine conditions. Journal of the Acoustical Society of America, 1997; 102: 3166